OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Education

The Fragility of Landmark Randomized Controlled Trials in the Plastic Surgery Literature

Benjamin H. Ormseth, BS*
Hassan ElHawary, MD, MSc†
Jeffrey E. Janis, MD, FACS*

Background: Randomized controlled trials (RCTs) are integral to the progress of evidenced-based medicine and help guide changes in the standards of care. Although results are traditionally evaluated according to their corresponding *P* value, the universal utility of this statistical metric has been called into question. The fragility index (FI) has been developed as an adjunct method to provide additional statistical perspective. In this study, we aimed to determine the fragility of 25 highly cited RCTs in the plastic surgery literature.

Methods: A PubMed search was used to identify the 25 highest cited RCTs with statistically significant dichotomous outcomes across 24 plastic surgery journals. Article characteristics were extracted, and the FI of each article was calculated. Additionally, Altmetric scores were determined for each study to determine article attention across internet platforms.

Results: The median FI score across included studies was 4 (2–7.5, interquartile range). The two highest FI scores were 208 and 58, respectively. Four studies (16%) had scores of 0 or 1. Three studies (12%) had scores of 2. All other studies (72%) had FI scores of 3 or higher. The median Altmetric score was 0 (0–3).

Conclusion: The FI can provide additional perspective on the robustness of study results, but like the *P* value, it should be interpreted in the greater context of other study elements. *(Plast Reconstr Surg Glob Open 2024; 12:e5352; doi: 10.1097/GOX.0000000000005352; Published online 17 January 2024.)*

INTRODUCTION

Randomized controlled trials (RCTs) are the medium of change when it comes to medical care. Although other forms of study typically provide the initial impetus forward, the stringent experimental design and randomization underlying RCTs enable investigators to best evaluate the efficacy of a given measure.[1] Most frequently, the results of RCTs are analyzed for statistical significance using a *P* value of less than 0.05. Although this benchmark is widely used throughout all of scientific literature, some have raised concerns over its universal utility.[2–4] Specifically, its validity can be heavily influenced by the sample size, event rate, and number lost to follow-up and, therefore, its value can be misinterpreted.[5]

From the *Department of Plastic and Reconstructive Surgery, The Ohio State University Wexner Medical Center, Columbus, Ohio; and †Division of Plastic and Reconstructive Surgery, McGill University Health Center, Montreal, Canada.

Received for publication May 19, 2023; accepted August 24, 2023.

DOI: 10.1097/GOX.0000000000005352

In an effort to add statistical perspective and provide an adjunct measure to *P* values, the fragility index (FI) has been developed to illustrate how "fragile" the findings of an RCT are.[6,7] It is defined as the minimum number of subjects in one arm of the trial requiring a change in outcome to lose statistical significance using the Fisher exact test. For example, for an FI of one, if one additional patient had a positive outcome who previously did not, the relative proportions of outcomes between study arms would no longer be significantly different. In recent years, this method has been applied throughout the surgical literature to better understand the fragility of RCTs in distinct subspecialties, including orthopedics, neurosurgery, colorectal surgery, and even plastic surgery.[8–12] However, these previous meta-analyses did not discuss the content or the impact of the articles they included, thus foregoing perspective on the true ramifications of their findings.

Considering that RCTs are responsible for evidence-based changes in patient standards of care, it stands to reason that the most impactful RCTs represent a unique

Disclosure statements are at the end of this article, following the correspondence information.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

echelon of true "game changers" in practice patterns. Thus, in this study, we aimed to focus our analysis of fragility on the most cited RCTs in plastic surgery to better assess the results of those studies with the most influence on the field.

METHODS

Search Method

A search was created to identify all articles indexed in PubMed through March of 2023 from 24 journals dedicated to the field of plastic and reconstructive surgery. The filter "randomized controlled trials" was applied to generate a list of studies. Additionally, a separate list was created by searching for the term "randomized trial" while excluding for reviews, systematic reviews, and meta-analyses. The combined list of PMIDs were then uploaded into iCite, a tool developed by the National Institutes of Health, to identify the total number of citations and the citations per year for each article.[13] From this list, the 25 most highly cited RCTs meeting study criteria were included. The literature search was conducted by authors B.O. and H.E., and any discrepancies between reviewers were resolved through consensus by a third author, J.E.J.

Study Criteria

Inclusion criteria were (1) at least one dichotomous outcome that was found to be statistically significant, (2) studies conducted on human participants, (3) studies pertaining to plastic and reconstructive surgery management, and (4) articles available in English. Exclusion criteria were (1) RCTs using a cluster or noninferiority design and (2) any other studies not meeting full inclusion criteria.

Data Collection

After identifying the 25 most highly cited RCTs meeting criteria, articles were individually reviewed, and the following variables were extracted: article title, authors, journal name, publication year, intervention, outcome, total sample size, sample size of the intervention group, number of events in the intervention group, sample size of the control group, number of events in the control group, *P* value, and sources of funding. Data extraction was verified by authors B.O. and H.E., and any discrepancies were resolved though consensus by J.E.J. If studies had a secondary outcome that was dichotomous but the primary outcome was not, the secondary outcome information was extracted. The number of times each article was cited was extracted from iCite. Additionally, article Altmetric scores were determined using the bookmark "Altmetric it!"[14] The Altmetric score is an alternative metric for measuring publication impact that represents the cumulative attention an article receives across internet platforms, including news articles, blog posts, Twitter, Facebook, Reddit, YouTube, and other web interfaces. The composite score is calculated by a weighted Altmetric algorithm according to source of article attention. Each article was then categorized according to best fit with one of the following section classifications used by *Plastic and Reconstructive Surgery–Global Open (PRS–GO)*: breast, burns, business, cosmetic, craniofacial/pediatric, education, gender-affirming surgery, global health, hand, peripheral nerve, reconstructive, research, technology, or wellness. Journal impact scores were determined using the Journal Citation Report 2021, Clarivate, when available.[15]

> **Takeaways**
>
> **Question:** How robust are the results of the most influential randomized controlled trials (RCTs) in plastic surgery as measured by the fragility index (FI)?
>
> **Findings:** Twenty-five of the highest cited RCTs were identified. The average citation count was 121.9 ± 24.7 SD across studies. The median FI was 4 (2-7.5), and 72% of studies had scores of 3 or higher.
>
> **Meaning:** The FI is a useful adjunct to more common statistical measures that can provide further perspective on the strength of results reported in RCTs.

FI Calculations and Data Analysis

An online calculator was used to determine the FI score of each study by iteratively converting one patient from a nonevent to positive event outcome and recalculating a two-sided Fisher exact test until the *P* value met or exceeded 0.05.[16] The final number of patients requiring conversion to lose significance was recorded as the FI. The FI was considered zero if the *P* value became nonsignificant upon recalculation with the Fisher exact test. Extracted variables were then summarized using descriptive statistics. Categorical variables were presented as frequencies and proportions. The Shapiro-Wilk test was used to determine the normality of continuous variables. Continuous variables were then presented as means and SDs or medians and interquartile ranges when appropriate. Data analyses and graphic production were performed using SPSS, version 27.0 (IBM Corp. Released 2020. IBM SPSS Statistics for Mac, Armonk, N.Y.: IBM Corp).

RESULTS

The 24 included journals and their corresponding impact factors are listed in Table 1. A total of 3334 articles were included in the search. (**See table, Supplemental Digital Content 1,** which displays the titles and associated article information of the top 25 most cited RCTs meeting criteria.[17–41] **http://links.lww.com/PRSGO/C823**.) The highest cited article was an RCT published in 2010 that investigated the efficacy of collagenase *Clostridium histolyticum* in treating Dupuytren disease by Gilpin et al (215 citations).[17] The study with the highest Altmetric score (22) was conducted by Guyuron et al, which demonstrated the utility of trigger point deactivation surgery in the treatment of migraine.[19] For articles with an Altmetric score greater than zero, the contributing sources of online attention are listed in Table 2.

The combined characteristics of the included studies are summarized in Table 3. The average citation count

Table 1. Journals Included in the PubMed Search

Journal	Impact Factor
Aesthetic Plastic Surgery	2.708
Aesthetic Surgery Journal	4.485
Annals of Plastic Surgery	1.763
Burns	2.609
Canadian Journal of Plastic Surgery	0.558
Cleft Palate-Craniofacial Journal	1.915
Clinics in Plastic Surgery	2.53
European Journal of Plastic Surgery	—
Indian Journal of Plastic Surgery	—
JAMA Facial Plastic Surgery	4.667
Journal of Burn Care & Research	1.819
Journal of Craniofacial Surgery	1.172
Journal of Craniomaxillofacial Surgery	3.192
Journal of Hand Surgery (European Volume)	2.206
Journal of Hand Surgery (American Volume)	2.342
Journal of Oral and Maxillofacial Surgery	2.136
Journal of Plastic Reconstructive & Aesthetic Surgery	3.022
Journal of Plastic Surgery and Hand Surgery	1.295
Journal of Reconstructive Microsurgery	2.329
Microsurgery	2.08
Ophthalmic Plastic and Reconstructive Surgery	2.011
Plastic and Reconstructive Surgery	5.169
Plastic and Reconstructive Surgery–Global Open	–
Scandinavian Journal of Plastic and Reconstructive Surgery	0.935

was 121.9 ± 24.7 SD, and the median Altmetric score was 0 (0–3, interquartile range). Only three studies had an Altmetric score above 10, and 56% of studies had an Altmetric score of 0. *Plastic and Reconstructive Surgery* had the highest percentage of RCTs included (48%), followed by *Journal of Hand Surgery* (*American Volume*, 20%), and *Burns* (16%). Only four of the included studies came from other journals. Articles pertained to the following *PRS–GO* section classifications: hand (28%), burns (20%), cosmetic (16%), breast (12%), craniofacial/pediatric (8%), peripheral nerve (8%), and reconstructive (8%). No articles were included that related to business, education, gender-affirming surgery, global health, research, technology, or wellness. Regarding the type of intervention used in the included RCTs, surgery (32%), therapeutic and cosmetic injections (24%), and wound treatments (20%) were the most commonly used methods. In terms of funding, only 32% of studies did not report any funding information, while 32% received industry funding; 16%, institutional funding; 4%, foundational funding; and 16% did not receive any funding at all. The median sample size of all included studies across control and intervention groups was 75 (48–136).

After calculating the FI of each study, the median FI was 4 (2–7.5) (Table 3). The two highest FI scores were 208 and 58, respectively. All other FI scores were 14 or lower. Four studies (16%) had scores of 0 or 1. Three studies (12%) had scores of 2. All other studies (72%) had FI scores of 3 or higher. The distribution and frequencies of corresponding FI scores can be seen in Figure 1, except for the two outlier values of 208 and 58. Of note, 36% of studies had a number lost to follow-up that was greater than their calculated FI scores.

DISCUSSION

RCTs provide scientific foundation to the progression of evidenced-based medicine and provide justification for changes in the current standards of care. Although their study design is inherently robust and well-suited to account for confounders, the utilization of the *P* value as an exclusive metric to evaluate results may lead to variable interpretations. In this study, we aimed to re-evaluate the most cited RCTs in the plastic surgery literature using the FI, a newly developed adjunct to the *P* value, to gain further perspective on the results of previous studies. The study was limited to the 25 highest cited RCTs meeting criteria to facilitate a focused analysis of those publications with the most impact on the field and to compare these results with findings from previous literature on the FI.

The key result of this study was the median FI score of 4 (2–7.5) across the 25 included RCTs. Previously, Chin et al conducted a meta-analysis of all RCTs in plastic surgery meeting criteria and found a median FI of 1 (0–4) across the 90 included studies.[12] When taken in the context of the current results, this demonstrates that narrowing the focus to the highest cited RCTs raises the median FI. Considering that citation count is a proxy for overall article impact, this ultimately suggests that the "game changing" RCTs in plastic surgery have comparatively more robust statistical differences in results between control and

Table 2. Breakdown of the Sources Included in the Composite Altmetric Score for Each Article

First Author	Pub Year	Altmetric Score	News Outlets	Blogs	Wikipedia Pages	Policy Source	Patents	Twitter	Facebook Page	Research Highlight Platform
Bahman Guyuron[19]	2009	22	1	1	4	0	0	4	0	0
Annet L. van Rijssen[18]	2012	13	0	0	1	1	1	9	1	1
Marie A Badalamente[25]	2007	12	0	0	1	1	11	0	0	0
Thomas C.M. Lundeberg[41]	1992	6	0	0	0	1	2	0	0	0
Bahman Guyuron[20]	2005	4	0	0	4	0	0	1	0	0
David Gilpin[17]	2010	3	0	0	0	0	2	0	0	0
Robert E. Marx[21]	1988	3	0	0	3	0	0	0	0	0
Marilyn E. Innes[27]	2001	3	0	0	0	0	2	0	0	0
Ruby Grewal[29]	2005	3	0	0	0	1	0	0	0	0
Paul Waymack[33]	2000	3	0	0	0	0	2	0	0	0
Daniel Murphy[35]	1995	3	0	0	0	1	0	0	0	0

Table 3. Summary of Combined Study Characteristics

Study Characteristic	N (%)
Citation count, average ± SD	121.9 ± 24.7
Altmetric Score, median (IQR)	0 (0–3)
Journal	
Plastic and Reconstructive Surgery	12 (48)
Journal of Hand Surgery (American Volume)	5 (20)
Burns	4 (16)
Other	4 (16)
Category	
Breast	3 (12)
Burns	5 (20)
Cosmetic	4 (16)
Craniofacial/pediatric	2 (8)
Hand	7 (28)
Peripheral nerve	2 (8)
Reconstructive	2 (8)
Other	0 (0)
Intervention	
Surgery	8 (32)
Injection, therapeutic or cosmetic	6 (24)
Wound treatment	5 (20)
Implant	2 (8)
Other	4 (16)
Funding	
Industry	8 (32)
Federal	0 (0)
Foundational	1 (4)
Institutional	4 (16)
None	4 (16)
Not reported	8 (32)
Total sample size, median (IQR)	75 (47.5–136)
FI, median (IQR)	4 (2–7.5)
FI compared with the no. lost to follow-up	
FI > No. lost to follow-up	14 (56)
FI < No. lost to follow-up	9 (36)
No. lost to follow-up not reported	2 (8)

intervention groups than the plastic surgery RCT literature at large, thus reinforcing confidence in their collective value. As comparators, meta-analyses on the fragility of other surgical literature have demonstrated a median FI of 3 for hand surgery, 1 (1–3) for esophageal surgery, 4.5 (1.5–10) for neurosurgery, 3 (1–10) for colorectal surgery, and 3 (1–7) for gynecological surgery, thus placing the current results near the higher end of ranges found in other surgical settings.[10,11,42–44] Additionally, a previous review of the nephrology literature found that 41% of 127 RCTs had numbers lost to follow-up greater than their corresponding FI scores, which is comparable to the value of 36% reported in the current study.[45]

Although looking at grouped FI scores gives an indication to the collective strength of the included studies, it is important to consider the results on an individual study level to gain further perspective. One benefit of the FI is that it can better illustrate just how significant a given P value less than 0.05 may be. In the study by Kane et al investigating Dysport (abobotulinumtoxinA) in the correction of glabellar lines, the comparative responder rates across study arms at 30 days was reported to be significant at P less than 0.001.[38] Although this communicates a meaningful difference, the corresponding FI score of 208 more intuitively demonstrates how large that difference is, as hundreds of participants would have to change outcomes to lose significance. Similarly, looking at the RCT investigating NexoBrid as an enzymatic debridement agent for burn wounds, the FI score of 58 provides a more tangible, impactful representation of the strength of the results than the reported P less than 0.0001.[31] Moreover, these FI scores have correlated to meaningful clinical developments, as well. Dysport has been integrated into the common cosmetic treatment of glabellar lines and has been demonstrated to yield high patient and physician satisfaction after multiple injections.[46] As of December 2022,

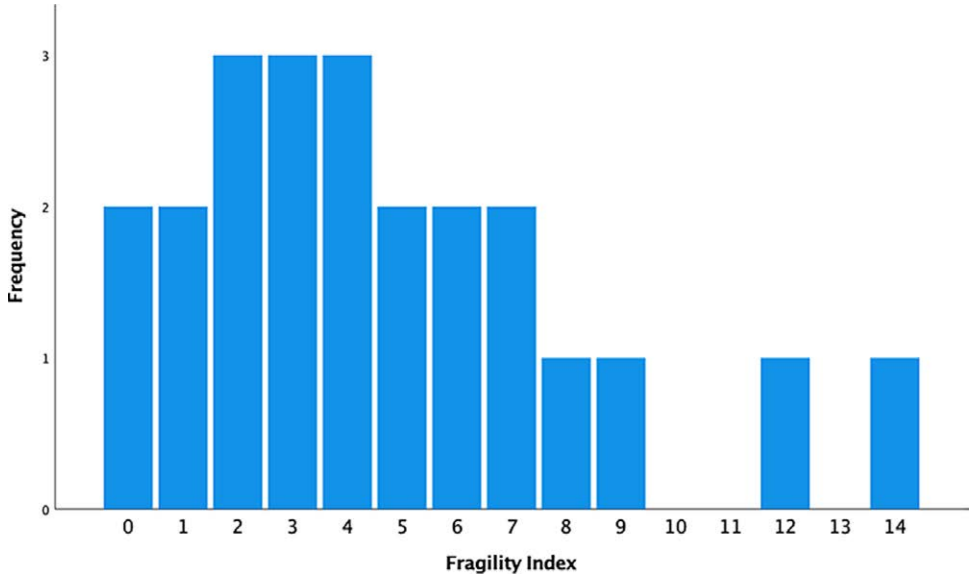

Fig. 1. Distribution of FI scores across 23 of 25 studies. Two studies with outlier scores (58 and 208) were omitted.

NexoBrid was officially approved by the Food & Drug Administration for treatment of severe thermal burns and is set to become commercially available in the second quarter of 2023.[47] Thus, the included studies with the highest FI scores have also corresponded with real-world, clinical changes over the long term.

Although outlier FI scores are simpler to digest, FI scores that increasingly approach zero become more difficult to interpret. An acknowledged weakness of the FI is that there is no consensus in the literature on how to interpret the magnitude of FI scores.[48] Without an established threshold in place, the audience is left to interpret what score is low enough to warrant hesitation about results.[49] Although an FI of 58 is clearly outside that window, and values closer to 10 may seem fairly robust, what about 5, 3, or even 1? How low is too low? Looking at the highest cited RCT included in the current study, Gilpin et al investigated the efficacy of injectable collagenase *Clostridium histolyticum* in treating Dupuytren disease[17] although 44% of patients in the intervention arm had a positive response at 30 days compared with only 5% of the control group ($P < 0.001$), the FI was 3. Depending on the reader, this FI score may depreciate an interpretation of results that should perhaps be more readily accepted. Additionally, studies with low FI scores have previously demonstrated highly impactful contributions to plastic surgery. In Dr. Guyuron's placebo-controlled sham surgery trial of trigger point deactivation surgery in 2009,[19] the FI score was only 2, yet this study has, in-large part, motivated continued investigations on the surgical treatment of migraine headaches, for which several large-scale meta-analyses now exist that demonstrate the robust efficacy and safety of these procedures.[50–53]

Another important consideration should be made for studies with an FI of 1 or 0. The FI was designed to quantify the robustness of a reported *P* value and add a layer of perspective beyond dichotomized significance.[6] Consequently, a score of one or zero can highlight tenuous significance according to a *P* value and warrant cautious interpretation. Considering the importance of RCTs in guiding management, this diligence is a useful adjunct that should be taken into consideration with other study elements[54]; however, it should also not invariably diminish the findings. Consider the study conducted by Lundeberg et al in 1992 on the efficacy of electrical nerve stimulation to improve the healing of diabetic ulcers, which reported a *P* value less than 0.05 and had an FI of zero.[41] Although this FI warrants careful reconsideration of the results, several additional RCTs on the same topic have followed in the wake of this study, and a recent 2022 meta-analysis on 10 RCTs concluded that electrical nerve stimulation may have benefit in this setting.[55] Thus, the original study by Lundeberg et al served as the impetus for 30 years of ongoing literature potentially progressing toward a useful clinical modality, and RCTs with similarly low FI scores may still serve a benefit in the context of greater research efforts.

In this study, we also determined Altmetric scores for the included articles as an additional tool to measure article exposure through mediums beyond citation counts. This method has previously been used in a study investigating the 50 most cited publications on blepharoplasty, which found a median Altmetric score of 1 (0–4.35).[56] The results of our study were similarly low with a median score of 0 (0–3). One plausible explanation for these low scores is that while plastic surgeons maintain an active online presence, this often manifests as marketing efforts,[57] and social media discussions of the literature may be more limited. Additionally, many of the included RCTs were published during the 1980s, 1990s, and early 2000s, which came before the widespread use of several of the platforms that Altmetric tracks, such as Twitter, Facebook, Reddit, and YouTube. In general, altmetrics represent a modern-day method to evaluate evidence-based medicine, and they offer advantages in gauging broader research impact beyond traditional metrics. Not only do they increase data diversity, but they may also provide faster, real-time feedback on article engagement, whereas subsequent article citations may take years to develop.[58] Nonetheless, altmetrics are still an evolving modality with concerns regarding data quality and a lack of rigorous evidence validating their utility. The ultimate role of altmetrics in the widespread evaluation of evidence-based medicine remains to be determined, and they may be more appropriate in the setting of more recently released literature.

LIMITATIONS

The FI traditionally has been limited to analyses of significantly different dichotomous outcomes. Although recent literature has begun to explore its utility in assessing continuous data,[59] time-to-event data,[60] and negative findings,[61] these settings have not been rigorously investigated, and our decision to focus on dichotomous outcomes with positive findings was further intended to preserve the homogeneity of results for meta-analysis. Additionally, because RCTs are powered to detect treatment effect on the primary outcome, applying the FI to secondary outcomes may be less robust.[62] Moreover, concerns have been raised on its close association with the *P* value and study sample size, and appropriate perspective requires further comparison with the number lost to follow-up of corresponding RCTs, although we have appropriately done so in the present study. As mentioned previously, the literature still lacks a consensus cutoff value to guide FI score interpretation. In addition, although less relevant to the current study that intentionally focused on high-impact RCTs, it should be noted that the FI cannot be applied to research with lower levels of evidence. Regarding other study limitations, although we aimed to be comprehensive with the journals included, there are other plastic surgery journals that may have yielded additional RCTs for consideration. Also, although our dual search strategy incorporating PubMed indexing of RCTs and studies included in the search term "randomized trial" was designed to be as inclusive as possible, it may be that further studies were not captured. Finally, we chose to use citation count to guide our selection of the most impactful RCTs in the plastic surgery literature. Although this is the most established and recognized proxy for article impact, it should be appreciated that this alone does not comprehensively define or identify the true clinical impact of any given article.

CONCLUSIONS

In our study, we demonstrated that the most cited RCTs in plastic surgery have a higher distribution of FI scores compared with RCTs in the field at large, suggesting that the most impactful RCTs are also more statistically robust. Although several statistical methods for evaluating RCT results exist, the optimal test is still under continued debate. The true impact of the FI is to translate a conceptual *P* value into a more tangible and intuitive output for audience understanding. As it continues to gain traction within the greater medical literature, it can provide value to future plastic surgery RCTs as a statistical adjunct that conveniently and effectively communicates the clinical meaning of study results.

Jeffrey E. Janis, MD, FACS
915 Olentangy River Road
Columbus, OH 43212
E-mail: jeffrey.janis@osumc.edu
Twitter: jjanismd
Instagram: jeffreyjanismd

DISCLOSURES

Dr. Janis receives royalties from Thieme and Springer Publishing. All the other authors have no financial interest to declare in relation to the content of this article.

REFERENCES

1. Guyatt G. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;268:2420.
2. Laber EB, Shedden K. Statistical significance and the dichotomization of evidence: the relevance of the ASA statement on statistical significance and p-values for statisticians. *J Am Stat Assoc*. 2017;112:902–904.
3. Vidgen B, Yasseri TP. Misunderstood and misused. *Front Phys*. 2016;4.
4. Sterne JAC, Davey Smith G. Sifting the evidence—what's wrong with significance tests? Another comment on the role of statistical methods. *BMJ*. 2001;322:226–231.
5. Thorlund K, Imberger G, Walsh M, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis—a simulation study. *PLoS One*. 2011;6:e25491.
6. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol*. 2014;67:622–628.
7. Liu Q, Chen H, Gao Y, et al. Robustness of significant dichotomous outcomes in randomized controlled trials in the treatment of patients with COVID-19: a systematic analysis. *Intensive Care Res*. 2023;3:38–49.
8. Megafu M, Megafu E. The fragility of statistical findings in distal radius fractures: a systematic review of randomized controlled trials. *Injury*. 2022;53:3352–3356.
9. Morris SC, Gowd AK, Agarwalla A, et al. Fragility of statistically significant findings from randomized clinical trials of surgical treatment of humeral shaft fractures: a systematic review. *World J Orthop*. 2022;13:825–836.
10. Nelms DW, Vargas HD, Bedi RS, et al. When the p value doesn't cut it: the fragility index applied to randomized controlled trials in colorectal surgery. *Dis Colon Rectum*. 2022;65:276–283.
11. Volovici V, Vogels VI, Dammers R, et al. Neurosurgical evidence and randomized trials: the fragility index. *World Neurosurg*. 2022;161:224–229.e14.
12. Chin B, Copeland A, Gallo L, et al. The fragility of statistically significant randomized controlled trials in plastic surgery. *Plast Reconstr Surg*. 2019;144:1238–1245.
13. ICite, Hutchins BI, Santangelo G. iCite Database Snapshots (NIH Open Citation Collection). Published online 2022.
14. Bookmarklet. Available at https://www.altmetric.com/products/free-tools/bookmarklet/. Accessed March 4, 2023.
15. Journal citation report. Clarivate; 2021. Available at https://access.clarivate.com/login?app=jcr&referrer=target%3Dhttps:%2F%2Fjcr.clarivate.com%2Fjcr%2Fhome&alternative=true&shibShireURL=https:%2F%2Flogin.incites.clarivate.com%2F%3FDestApp%3DIC2JCR%26amp;auth%3DShibboleth&shibReturnURL=https:%2F%2Flogin.incites.clarivate.com%2F. Accessed February 15, 2023.
16. Kane S. Fragility index calculator. Available at https://clincalc.com/Stats/FragilityIndex.aspx. Accessed February 15, 2023.
17. Gilpin D, Coleman S, Hall S, et al. Injectable collagenase *Clostridium histolyticum*: a new nonsurgical treatment for Dupuytren's disease. *J Hand Surg*. 2010;35:2027–38.e1.
18. van Rijssen AL, Ter Linden H, Werker PMN. Five-year results of a randomized clinical trial on treatment in Dupuytren's disease: percutaneous needle fasciotomy versus limited fasciectomy. *Plast Reconstr Surg*. 2012;129:469–477.
19. Guyuron B, Reed D, Kriegler JS, et al. A placebo-controlled surgical trial of the treatment of migraine headaches. *Plast Reconstr Surg*. 2009;124:461–468.
20. Guyuron B, Kriegler JS, Davis J, et al. Comprehensive surgical treatment of migraine headaches. *Plast Reconstr Surg*. 2005;115:1–9.
21. Marx RE, Morales MJ. Morbidity from bone harvest in major jaw reconstruction: a randomized trial comparing the lateral anterior and posterior approaches to the ilium. *J Oral Maxillofac Surg*. 1988;46:196–203.
22. Carruthers JD, Lowe NJ, Menter MA, et al; Botox Glabellar Lines II Study Group. Double-blind, placebo-controlled study of the safety and efficacy of botulinum toxin type A for patients with glabellar lines. *Plast Reconstr Surg*. 2003;112:1089–1098.
23. Atkinson JAM, McKenna KT, Barnett AG, et al. A randomized, controlled trial to determine the efficacy of paper tape in preventing hypertrophic scar formation in surgical incisions that traverse Langer's skin tension lines. *Plast Reconstr Surg*. 2005;116:1648–56; discussion 1657–1658.
24. Worsaae N, Thorn JJ. Surgical versus nonsurgical treatment of unilateral dislocated low subcondylar fractures: a clinical study of 52 cases. *J Oral Maxillofac Surg*. 1994;52:353–60; discussion 360–361.
25. Badalamente MA, Hurst LC. Efficacy and safety of injectable mixed collagenase subtypes in the treatment of Dupuytren's contracture. *J Hand Surg Am*. 2007;32:767–774.
26. Still J, Glat P, Silverstein P, et al. The use of a collagen sponge/living cell composite material to treat donor sites in burn patients. *Burns*. 2003;29:837–841.
27. Innes ME, Umraw N, Fish JS, et al. The use of silver coated dressings on donor site wounds: a prospective, controlled matched pair study. *Burns*. 2001;27:621–627.
28. Asplund O. Capsular contracture in silicone gel and saline-filled breast implants after reconstruction. *Plast Reconstr Surg*. 1984;73:270–275.
29. Grewal R, Perey B, Wilmink M, et al. A randomized prospective study on the treatment of intra-articular distal radius fractures: open reduction and internal fixation with dorsal plating versus

mini open reduction, percutaneous fixation, and external fixation. *J Hand Surg Am.* 2005;30:764–772.

30. Burkhardt BR, Eades E. The effect of Biocell texturing and povidone-iodine irrigation on capsular contracture around saline-inflatable breast implants. *Plast Reconstr Surg.* 1995;96:1317–1325.
31. Rosenberg L, Krieger Y, Bogdanov-Berezovski A, et al. A novel rapid and selective enzymatic debridement agent for burn wound management: a multi-center RCT. *Burns.* 2014;40:466–474.
32. Sproat JE, Dalcin A, Weitauer N, et al. Hypertrophic sternal scars: silicone gel sheet versus Kenalog injection treatment. *Plast Reconstr Surg.* 1992;90:988–992.
33. Waymack P, Duff RG, Sabolinski M. The effect of a tissue engineered bilayered living skin analog, over meshed split-thickness autografts on the healing of excised burn wounds. The Apligraf Burn Study Group. *Burns.* 2000;26:609–619.
34. Wilhelmi BJ, Blackwell SJ, Miller JH, et al. Do not use epinephrine in digital blocks: myth or truth? *Plast Reconstr Surg.* 2001;107:393–397.
35. Murphy D, Failla JM, Koniuch MP. Steroid versus placebo injection for trigger finger. *J Hand Surg Am.* 1995;20:628–631.
36. Fearon JA, Weinthal J. The use of recombinant erythropoietin in the reduction of blood transfusion rates in craniosynostosis repair in infants and children. *Plast Reconstr Surg.* 2002;109:2190–2196.
37. Waikakul S, Wongtragul S, Vanadurongwan V. Restoration of elbow flexion in brachial plexus avulsion injury: comparing spinal accessory nerve transfer with intercostal nerve transfer. *J Hand Surg Am.* 1999;24:571–577.
38. Kane MAC, Brandt F, Rohrich RJ, et al; Reloxin Investigational Group. Evaluation of variable-dose treatment with a new U.S. botulinum toxin type A (Dysport) for correction of moderate to severe glabellar lines: results from a phase III, randomized, double-blind, placebo-controlled study. *Plast Reconstr Surg.* 2009;124:1619–1629.
39. Cruz-Korchin N, Korchin L. Vertical versus Wise pattern breast reduction: patient satisfaction, revision rates, and complications. *Plast Reconstr Surg.* 2003;112:1573–1578; discussion 1579–1581.
40. Talon MD, Woodson LC, Sherwood ER, et al. Intranasal dexmedetomidine premedication is comparable with midazolam in burn children undergoing reconstructive surgery. *J Burn Care Res.* 2009;30:599–605.
41. Lundeberg TC, Eriksson SV, Malm M. Electrical nerve stimulation improves healing of diabetic ulcers. *Ann Plast Surg.* 1992;29:328–331.
42. Lee Y, Samarasinghe Y, Javidan A, et al. The fragility of significant results from randomized controlled trials in esophageal surgeries. *Esophagus.* 2023;20:195–204.
43. Pascoal E, Liu M, Lin L, et al. The fragility of statistically significant results in gynaecologic surgery: a systematic review. *J Obstet Gynaecol Can.* 2022;44:508–514.
44. Ruzbarsky JJ, Khormaee S, Daluiski A. The fragility index in hand surgery randomized controlled trials. *J oHand Surg.* 2019;44:698.e1–698.e7.
45. Shochet LR, Kerr PG, Polkinghorne KR. The fragility of significant results underscores the need of larger randomized controlled trials in nephrology. *Kidney Int.* 2017;92:1469–1475.
46. Gubanova E, Haddad Tabet M, Bergerova Y, et al. Assessment of subject and physician satisfaction after long-term treatment of glabellar lines with abobotulinumtoxinA (Dysport/Azzalure): primary results of the appeal noninterventional study. *Aesth Plast Surg.* 2018;42:1672–1680.
47. MediWound Announces FDA Approval of NexoBrid® for the Treatment of Severe Thermal Burns in Adults. GlobeNewswire; 2022. Available at https://www.globenewswire.com/news-release/2022/12/29/2580838/30505/en/MediWound-Announces-FDA-Approval-of-NexoBrid-for-the-Treatment-of-Severe-Thermal-Burns-in-Adults.html. Accessed March 1, 2023.
48. Lin L, Xing A, Chu H, et al. Assessing the robustness of results from clinical trials and meta-analyses with the fragility index. *Am J Obstet Gynecol.* 2023;228:276–282.
49. Carter RE, McKie PM, Storlie CB. The fragility index: a *P*-value in sheep's clothing? *Eur Heart J.* 2016;38:ehw495.
50. Hatef DA, Gutowski KA, Culbertson GR, et al. A comprehensive review of surgical treatment of migraine surgery safety and efficacy. *Plast Reconstr Surg.* 2020;146:187e–195e.
51. Janis JE, Barker JC, Javadi C, et al. A review of current evidence in the surgical treatment of migraine headaches. *Plast Reconstr Surg.* 2014;134:131S–141S.
52. ElHawary H, Barone N, Baradaran A, et al. Efficacy and safety of migraine surgery: a systematic review and meta-analysis of outcomes and complication rates. *Ann Surg.* 2022;275:e315–e323.
53. Huayllani MT, Janis JE. Migraine surgery and determination of success over time by trigger site: a systematic review of the literature. *Plast Reconstr Surg.* 2022;151:120e–135e.
54. Acuna SA, Sue-Chue-Lam C, Dossa F. The fragility index— *P* values reimagined, flaws and all. *JAMA Surg.* 2019;154:674.
55. Zheng Y, Du X, Yin L, et al. Effect of electrical stimulation on patients with diabetes-related ulcers: a systematic review and meta-analysis. *BMC Endocr Disord.* 2022;22:112.
56. Evereklioglu C, Sener H, Horozoglu F. Top 50 most-cited publications on blepharoplasty surgery between 2015 and 2022: from a current altmetric perspective of research impact. *Indian J Plast Surg.* 2023;56:118–123.
57. Shauly O, Marxen T, Goel P, et al. The new era of marketing in plastic surgery: a systematic review and algorithm of social media and digital marketing. *Aesthet Surg J Open Forum.* 2023;5:ojad024.
58. Bornmann L. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *J Inform.* 2014;8:895–903.
59. Caldwell JME, Youssefzadeh K, Limpisvasti O. A method for calculating the fragility index of continuous outcomes. *J Clin Epidemiol.* 2021;136:20–25.
60. Bomze D, Asher N, Hasan Ali O, et al. Survival-inferred fragility index of phase 3 clinical trials evaluating immune checkpoint inhibitors. *JAMA Netw Open.* 2020;3:e2017675.
61. Li A, Javidan AP, Liu E, et al. Assessing the robustness of negative vascular surgery randomized controlled trials using their reverse fragility index. *J Vasc Surg.* 2022;78:253–259.e11.
62. Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg.* 2019;154:74–79.